

# Hybrid Recommendation System

Jiarui Li  
Sijia Wei  
Tim Zhang  
Junyun Huang

## Methods Overall

### Embedding Model:

- Trained on Microsoft's MIND dataset using BERT-base-uncased with a Masked Language Modeling (MLM) approach.
- Preprocessed data by combining multiple fields into "news info" and dropping entries exceeding 512 characters.
- Utilized Transformers API for handling training, masking, and model updates.

### Content-Based Filtering (CBF):

- Generated recommendations by analyzing user reading history with transformer-based semantic embeddings.
- Limited comparisons to articles within the same subcategory for computational efficiency.
- Ranked articles using cosine similarity to ensure personalized, contextually relevant recommendations.

### Collaborative Filtering (CF):

- Grouped users based on interaction patterns using clustering techniques.
- Recommended popular articles from peer groups to balance exploration and familiarity.
- Enhanced diversity by introducing varied yet popular content, adapting to unconventional user preferences.

### Collaborative Filtering Testing

#### Sample of user's view history:

```
News ID: N55189, Category: tv, Title: 'Wheel Of Fortune' Guest Delivers Hilarious, Off The Rails Introduction
News ID: N42782, Category: sports, Title: Three takeaways from Yankees' ALCS Game 5 victory over the Astros
News ID: N34694, Category: tv, Title: Rosie O'Donnell: Barbara Walters Isn't 'Up to Speaking to People' Right Now
News ID: N45794, Category: news, Title: Four flight attendants were arrested in Miami's airport after bringing in thousands in cash, police
News ID: N18445, Category: sports, Title: Michigan sends breakup tweet to Notre Dame as series goes on hold
News ID: N63302, Category: lifestyle, Title: This Wedding Photo of a Canine Best Man Captures Just How Deep a Dog's Love Truly Is
News ID: N18414, Category: movies, Title: Robert Evans, 'Chinatown' Producer and Paramount Chief, Dies at 89
News ID: N19347, Category: news, Title: Former US Senator Kay Hagan dead at 66
News ID: N31801, Category: news, Title: Joe Biden reportedly denied Communion at a South Carolina church because of his stance on abortion
```

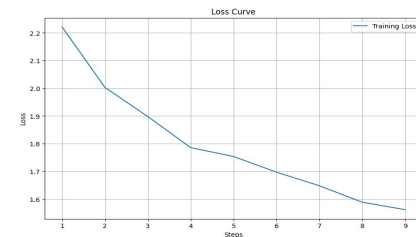
#### Recommendation results from Collaborative Filtering:

```
News ID: N18445, Category: sports, Title: Michigan sends breakup tweet to Notre Dame as series goes on hold
News ID: N34694, Category: tv, Title: Rosie O'Donnell: Barbara Walters Isn't 'Up to Speaking to People' Right Now
News ID: N42782, Category: sports, Title: Three takeaways from Yankees' ALCS Game 5 victory over the Astros
News ID: N45794, Category: news, Title: Four flight attendants were arrested in Miami's airport after bringing in thousands in ca:
News ID: N55189, Category: tv, Title: 'Wheel Of Fortune' Guest Delivers Hilarious, Off The Rails Introduction
```

### Content-Based Filtering Testing

```
11
Article ID: N30924
Title: The Rock's Gnarly PalM Is a Testament to Life Without Lifting Gloves
Top 5 Most Relevant Articles:
ID: N44111, Title: 14 Celebs Over 50 Who Are In The Best Shape Of Their Lives, Similarity: 0.8807
ID: N9140, Title: 85-Year-Old Runner Who Keeps Racking Up Medals: 'It's My Job to Get People Moving', Similarity: 0.8735
ID: N58885, Title: How This Guy Built Muscle and Got Shredded in His 40s on a Plant-Based Diet, Similarity: 0.8735
ID: N53700, Title: Ellen DeGeneres Likes to Do This Quick Abs Workout Before Hitting the Stage for Her Talk Show, Simila
ID: N44005, Title: How David Boreanaz Stays 'Seal Team' Fit at Age 50, Similarity: 0.8691
```

```
100%|██████████| 1/1 [00:01<00:00, 1.33s/it]12
Article ID: N6778
Title: A part of Windsor Castle that's been closed to the public for over 150 years just reopened and it's been comple
Top 5 Most Relevant Articles:
ID: N8071, Title: The Brands Queen Elizabeth, Prince Charles, and Prince Philip Swear By, Similarity: 0.8456
ID: N19818, Title: Kate Middleton's Best Hairstyles Through the Years, Similarity: 0.8450
ID: N1810, Title: Every outfit Duchess Kate has worn in 2019, Similarity: 0.8392
ID: N53921, Title: 25 Photos of the Royal Family at Balmoral Castle, Queen Elizabeth's Favorite Home, Similarity: 0.8344
ID: N63495, Title: Prince George's Royal Life in Photos, Similarity: 0.8223
```



Step	Training Loss	Validation Loss
1000	2.221500	1.876811
2000	2.003200	1.775177
3000	1.898000	1.693791
4000	1.785800	1.637199
5000	1.753900	1.590190
6000	1.697300	1.546615
7000	1.648300	1.516157
8000	1.588900	1.488654
9000	1.562000	1.458267